DOCUMENT RESUME

ED 093 939 95 TM 003 757

AUTHOR Klein, Roger D.; And Others
TITLE The Measurement and Development of "Accuracy" in Academic Performance.
INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE 73
NOTE 18p.

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Academic Achievement; Behavioral Science Research; Behavior Change; Measurement Techniques; *Reinforcement; Response Style (Tests); *Scoring; Testing
IDENTIFIERS *Accuracy Measures

ABSTRACT

In behavioral studies of academic performance, accuracy has usually been defined as the number of items correct divided by the number of items assigned. One previous study used an alternative definition--the number of items correct divided by the number of items attempted. It is suggested here that while both measures are useful indices of behavior, they need to be carefully distinguished. Two behavior modification experiments are presented which illustrate the usefulness of reporting both measures of accuracy. It was shown that during the second baseline stage of each study, accuracy based on items assigned decreased, while accuracy based on items attempted remained high. Suggestions are offered to explain this phenomenon. (Author)

ED 093939

THE MEASUREMENT AND DEVELOPMENT OF "ACCURACY" IN ACADEMIC PERFORMANCE

Roger D. Klein, Linda Berrett, and Robert Grant

Learning Research and Development Center
University of Pittsburgh

1973

TM 003 757

The research reported herein was supported by the Learning Research and Development Center, supported in part by funds from the National Institute of Education (NIE), United States Department of Health, Education, and Welfare. The opinions expressed do not necessarily reflect the position or policy of NIE and no official endorsement should be inferred.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Abstract

In behavioral studies of academic performance, accuracy has usually been defined as the number of items correct divided by the number of items assigned. One previous study used an alternative definition--the number of items correct divided by the number of items attempted. It is suggested here that while both measures are useful indices of behavior, they need to be carefully distinguished. Two behavior modificiation experiments are presented which illustrate the usefulness of reporting both measures of accuracy. It was shown that during the second baseline stage of each study, accuracy based on items assigned decreased, while accuracy based on items attempted remained high. Suggestions are offered to explain this phenomenon.

THE MEASUREMENT AND DEVELOPMENT OF "ACCURACY" IN ACADEMIC PERFORMANCE

Roger D. Klein
Learning Research and Development Center
University of Pittsburgh

Linda Berrett
Turner Elementary School
Wilkinsburg School District

and

Robert Grant
Finley Junior High School
Ringgold School District

Numerous classroom management studies have demonstrated that on-task or study behavior can be effectively increased (Bushell, Wrobel, & Michaelis, 1968; Hall, Lund, & Jackson, 1968). Recently, there has also been an interest in modifying other aspects of academic performance such as assignment completion (Klein & Mechelli, 1973; McLaughlin & Malaby, 1972), performance rate (Kirby & Shields, 1972), and performance accuracy (Conlon, Hall, & Hanley, 1972; Ferritor, Buckholdt, Hamblin, & Smith, 1972; Lovitt, Guppy, & Blattner, 1969; Sulzer, Hunt, Ashby, Komarski, & Krams, 1971).

With the exception of the Ferritor et al. study, studies reporting accuracy data usually define accuracy as the number of items correct divided by the number of items assigned. Ferritor et al., however, used an alternative definition by substituting "attempted" for "assigned," thus making their definition of accuracy the number of items correct divided by the number of items attempted.

Both types of accuracy can be useful when summarizing a given set of data. However, in applied academic investigations it may be important to report both accuracy measures, and distinguish between them, because under certain conditions these measures can show wide discrepancies and can thus produce different interpretations of the results. Two illustrations from a previously published investigation further this position.

Figure 1 reproduces the data on arithmetic performance in the first of two experiments reported by Ferritor et al. Students were assigned 100 arithmetic items per day throughout all phases of the investigation. The data in Figure 1 provide information on the median number of correct items and the median percent of items correct. The latter term was referred to as "accuracy" by Ferritor et al., and was derived from the formula: the number of items correct divided by the number of items attempted. As seen in Figure 1, accuracy was substantially higher during phases C_1 and B_2, than during the baseline phase. However, the median number of items correct remained stable during these phases. A rather substantial increase in accuracy was reported, even though there was virtually no increase in the number of items correct. Because the number of items assigned was constant, the increase in accuracy can only result from students attempting fewer items. If the first definition of accuracy presented in this paper, based on items assigned is applied to the Ferritor et al. data, no marked increase in accuracy is apparent.

Figure 2 presents the data on arithmetic performance from the second experiment reported by Ferritor et al. As can be seen, an increase in accuracy, in this case from baseline to phase C_1, was not accompanied by a corresponding acceleration in the number of items correct.

The reasons for such findings are relatively simple. Assuming a constant number of items is assigned, the formula based on items assigned is affected by only one variable--items correct. An increase in accuracy can only occur if the number of items correct increases. On the other hand, the accuracy formula based on items attempted is affected by two variables--items correct and items attempted. Thus, as seen in

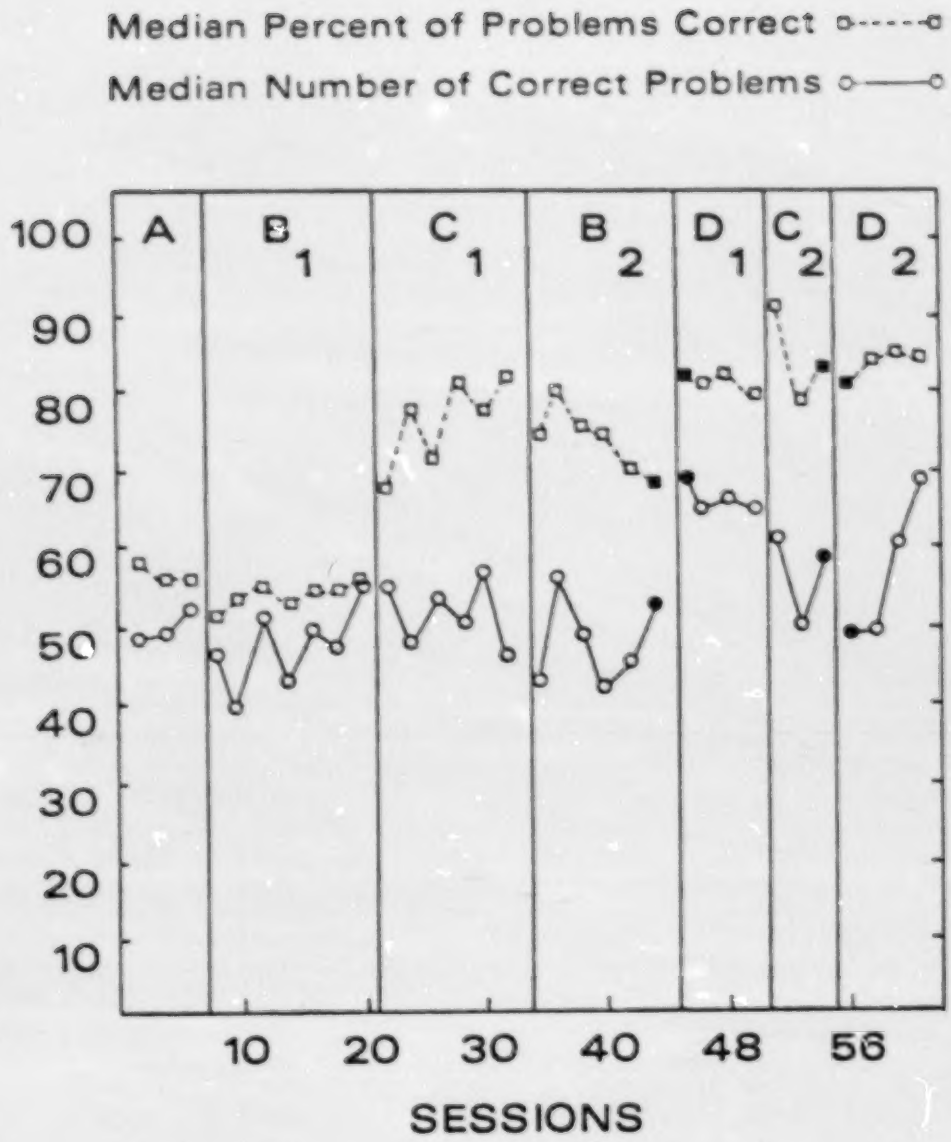

Figure 1. Median number of problems worked correctly and median percent of problems worked correctly by 14 third-grade children during a time when the children worked on 100 arithmetic computation problems. Individual percents were calculated by dividing the number correct by the number attempted. After the baseline condition (A), the children went through conditions in which reinforcement was contingent upon attending behavior (B), arithmetic performance (C) and a combination of arithmetic performance and attending behavior (D). Filled points are for single sessions; all others are combined data for two sessions. (From Ferritor et al., 1972, p. 12.)

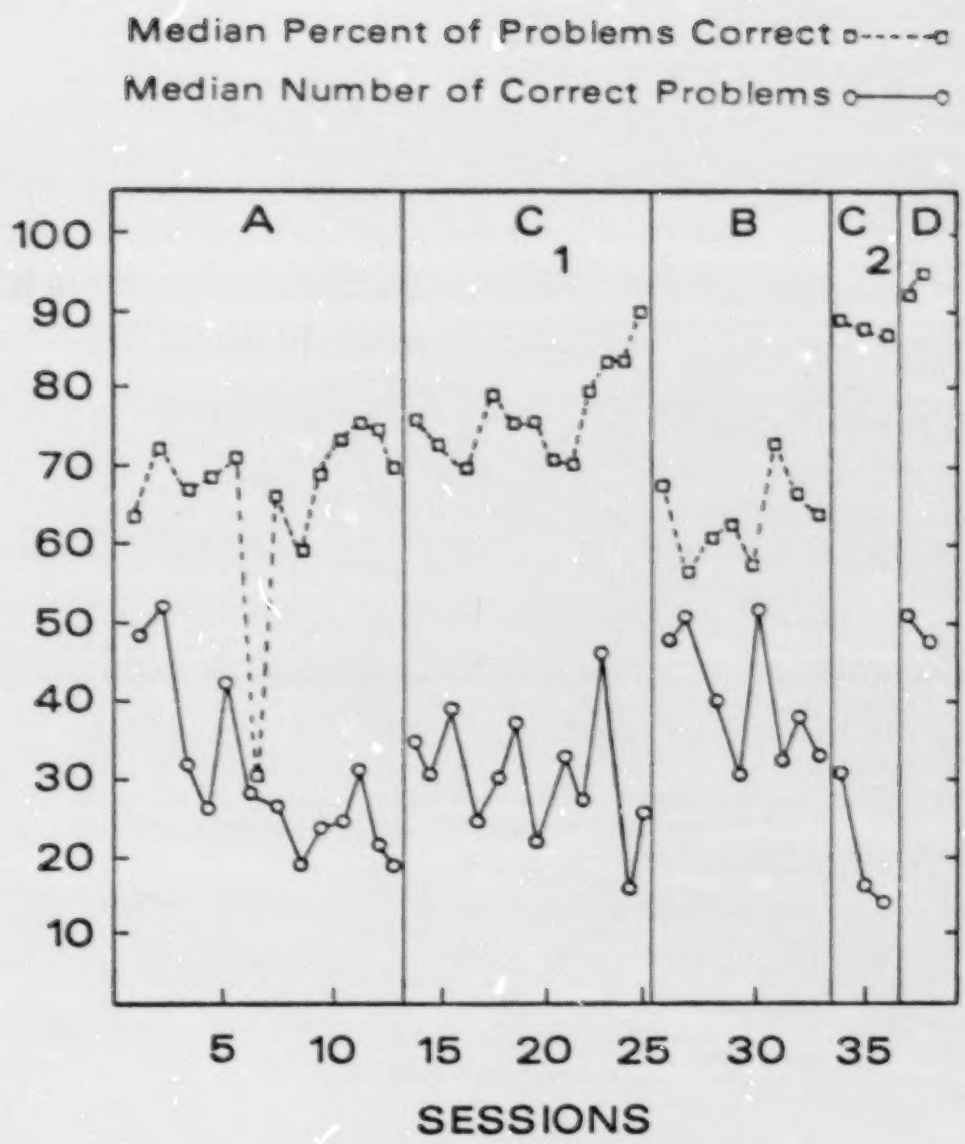

Figure 2. Median number of arithmetic problems worked correctly and median percent worked correctly for a group of nine third-graders working 100 computational problems. Individual percentages were calculated by dividing the number correct by the number attempted. After the baseline condition (A), the children went through conditions in which reinforcement was contingent upon arithmetic performance (C), attending behavior (B), and finally a combination of arithmetic performance and attending behavior (D). (From Ferritor et al., 1972, p. 15.)

the study by Ferritor et al., accuracy increased even as the number of items correct remained the same.

In addition, accuracy can also increase if the number of items correct decreases, providing there is a greater decrease in items attempted. This latter phenomenon is also illustrated in Figure 2. From stage B to C_2 a sharp increase was seen in accuracy, while the median number of items correct decreased. Again, because items assigned were held constant, the only explanation for the rise in the level of accuracy is that there was a greater decrease in items attempted than in items correct.

In many cases, accuracy based on items assigned is perhaps preferred, as a measure, over the items attempted formula. This may be true because in the extreme case the items attempted formula can indicate perfect accuracy, if the subject attempts only one item and performs it correctly. However, this does not imply that the items attempted measure is useless. Providing that a sufficiently high number of problems are attempted, some interesting phenomena can be studied.

For instance, many published classroom behavior modification studies include a reversal design. Some of these studies (e.g., Conlon et al., 1972) show that during a reversal period performance decreased, but did not always approach the original baseline level. It may be assumed that the failure to return to the original baseline performance is due to a number of factors (e.g., resistance to extinction, length of the reversal period). There is generally, however, no attempt made to determine the cause. The present two investigations suggest that partial reversals in studies involving accuracy data may be due in some instances to a resistance to extinction of a specific factor--accuracy based on items attempted.

The two studies in this paper are presented in order to illustrate the differential effects of an experimental reversal upon the two measures of accuracy, and to recommend that data on items attempted always be reported. In addition, some suggestions are offered as to why accuracy based on items attempted may be resistant to extinction.

The research reported here was conducted entirely by regular classroom teachers (second and third authors) who selected the students, designed the modification program, and recorded the results. The teachers were enrolled in an Educational Psychology course offered by the first author, and they received regular weekly feedback during class sessions. Completion of the project fulfilled part of the requirements for completion of the course. The present studies, therefore, further support previous work (e.g., Hall, 1971) which demonstrated that teachers can be easily trained to design and conduct behavior management investigations.

Experiment I

Method

Subject and Setting. Vic was a 14-year-old sixth grade student in a regular public school. He was approximately two years older than most of the students in his class because he entered first grade late, and because he repeated third grade due to academic difficulties. Vic's arithmetic skills were measured at the third grade level (Stanford Achievement Test) and, thus, the teacher attempted to individualize his assignments at that level.

Procedure. The math period extended from 10:10 a.m. until 10:45 a.m., five days a week. While the remaining students received group instruction, Vic was assigned a worksheet for Monday through

Thursday with 30 arithmetic problems. The problems were designed by the teacher (second author) and consisted of equal numbers of two-digit addition and subtraction items, emphasizing carrying and borrowing, and basic multiplication and division items. The multiplication items consisted of one-digit numbers and the division items had two-digit dividends and one-digit divisors.

The worksheet tested arithmetic skills that had been previously learned and no attempt was made to teach new skills during the experiment. No problem was assigned more than once. On Fridays, in place of the worksheet, Vic spent the full time (35 minutes) with the teacher going over problems with which he had experienced difficulty.

The experiment consisted of a four-phase ABAB reversal design and lasted for a total of 27 school days.

Baseline$_1$. During this phase, the teacher presented Vic with the worksheet and told him that he had exactly 35 minutes in which to complete it. At the end of the period, the worksheet was corrected by the teacher and returned to Vic with the number of items correct indicated on the top of the sheet. This phase lasted eight days.

Reinforcement$_1$. In talking with Vic, the teacher determined that Vic enjoyed erasing and washing the class blackboards. A contingency contract was designed in which it was agreed that on those days when Vic correctly completed 20 of the 30 problems assigned, he would be allowed to care for the boards. Vic was permitted to engage in this activity after school, between 3:00 p.m. and 3:20 p.m., while the teacher was present. On days when Vic correctly completed 27 of the 30 problems assigned, in addition to caring for the blackboards, he was permitted to select and keep one piece of construction paper from the teacher's paper supplies. This phase lasted eleven days.

Baseline $_2$. The teacher explained to Vic that the contract was no longer in effect. Baseline$_1$ conditions were reinstated. This phase remained in effect for four days.

Reinforcement$_2$. The contract described in reinforcement$_1$ was reinstated. The phase lasted for four days.

Results

Figure 3 presents Vic's accuracy scores on a daily basis. When accuracy was calculated by the formula using items assigned, it can be seen that accuracy averaged 27 percent during baseline$_1$, increased to 77 percent during reinforcement$_1$, fell to 65.5 percent during baseline$_2$ and rose again to 80.3 percent in reinforcement$_2$. When accuracy was calculated by the items attempted method, the average results indicate 56 percent during baseline$_1$, 79.4 percent during reinforcement$_1$, 82.5 percent during baseline$_2$, and 86.3 percent during reinforcement$_2$. The major finding was that in the items attempted measure, there was a slight increase rather than a reversal apparent during baseline$_2$. This contrasts with the 12 percent mean decrease for the same phase in the items assigned data.

The reasons for the difference in accuracy measures can be seen in Table 1, which presents mean data on the number of items attempted and the number correct. While the mean number of items correct decreased from reinforcement$_1$ to baseline$_2$ by 3.5, the mean number attempted decreased by 6.2. The reduction in mean number correct caused the accuracy score based on items assigned to decrease. However, the more marked reduction in mean items attempted resulted in the increase in accuracy based on items attempted. In other words, during baseline$_2$, Vic attempted 20 percent fewer items than during

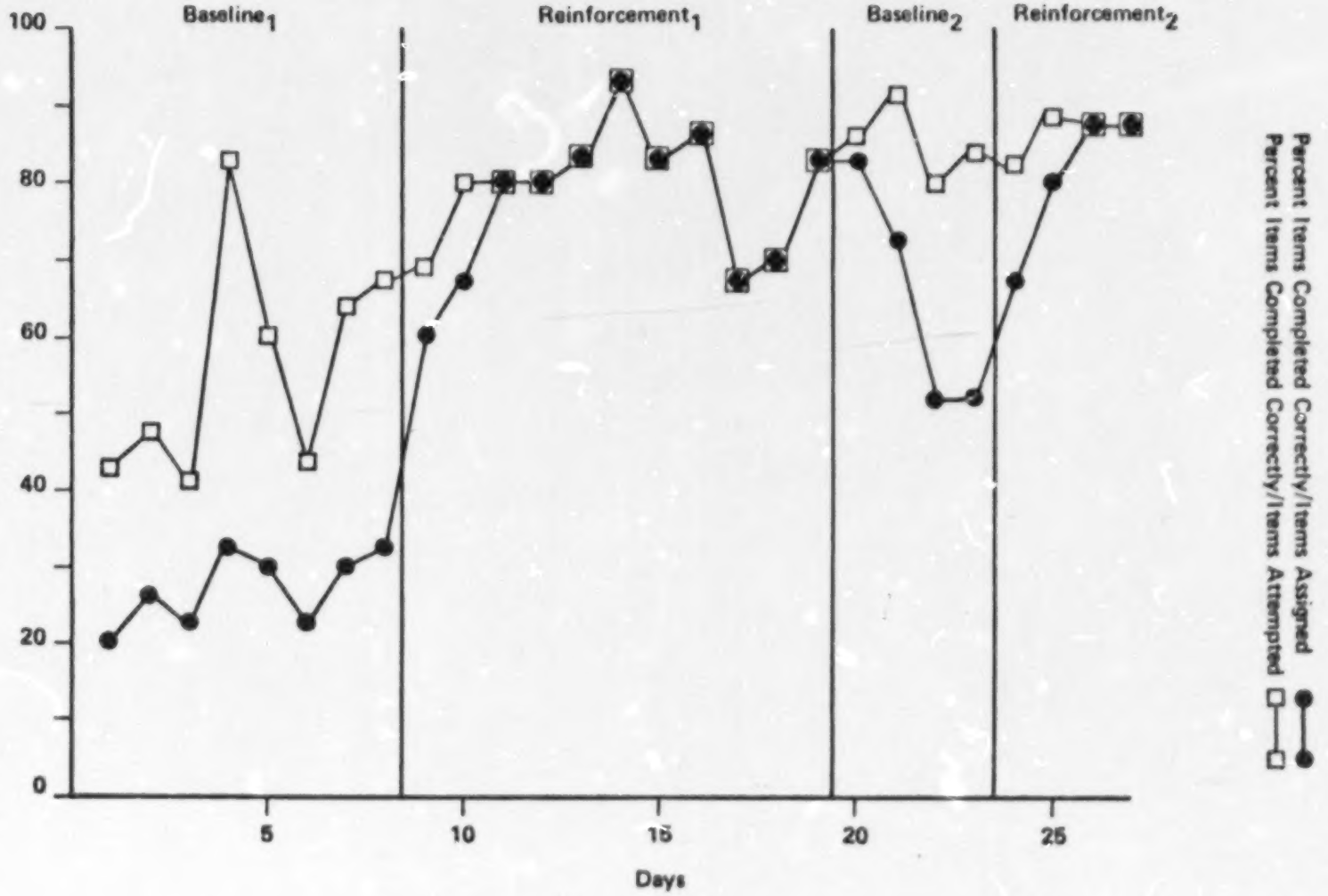

Figure 3. Measures of accuracy based on items assigned and items attempted for each phase of Experiment I.

TABLE 1

Mean Data on Percent of Problems Attempted and Number Correct per Day for Each Phase of Experimant 1 (Vic)

Phase	Mean Number of Problems Attempted per Day in Each Phase	Mean Number Correct per Day in Each Phase
$Baseline_1$	14.7	8.3
$Reinforcement_1$	29.3	23.3
$Baseline_2$	23.1	19.8
$Reinforcement_2$	27.8	24.0

$reinforcement_1$, but of those attempted, most were answered correctly. It should, however, be carefully noted that in spite of the reduction in number of items attempted between $reinforcement_1$ and $baseline_2$, the $baseline_2$ daily average of items attemnted was still 8.4 above that in $baseline_1$. Had the $baseline_2$ daily average of items attempted dropped to a very low level, then an accuracy score based on items attempted would have been inflated and would have presented a different picture of performance.

Experiment II

The second study was concerned with mathematics homework behavior. The data are used as partial support for the first study since daily scores were not available and the data presented are averages for each phase of the study.

Method

Subject and Setting. Susan was a 14 year old in a class of 30 ninth grade algebra students in a public junior high school. She rarely turned in homework assignments, and when she did, her performance on them was poor.

Procedure. All students were assigned identical homework tasks five days per week. Assignments were taken directly from the algebra textbook being used in class and the items progressed along a continuum of increasing difficulty throughout the study.

The experiment consisted of a four-phase ABAB reversal design lasting 19 days.

The number of problems assigned varied each day, but the averages were fairly constant over the four phases, 16.7, 14.5, 15, and 16.2 problems per day, respectively. The treatments consisted of:

Baseline$_1$. The teacher collected, marked, and returned all students' homework assignments but no comments were written on the papers. This phase lasted eight days.

Reinforcement$_1$. Baseline procedures remained in effect for all students except Susan. On Susan's paper, positive comments were written such as "Good work," "Good improvement," "Fine paper." This phase lasted four days.

Baseline$_2$. Baseline$_1$ conditions were reestablished and no comments were written. This phase lasted four days.

Reinforcement$_2$. Reinforcement$_1$ conditions were reinstated during a three-day phase.

During each phase of the experiment, a quiz was given in class in order to assess mastery of the assigned work. The items on the quiz were taken from the class textbook and were similar but not identical to the homework problems.

Results

Table 2 presents average data on Susan's algebra performance. The percent of items attempted and both types of accuracy measures are shown. Her test scores for each phase of the study are also presented.

TABLE 2

Mean Data on Both Measures of Accuracy, Percent of Items Attempted, and Test Scores During Each Phase of Experiment II (Susan)

Phase	Mean Percent Accuracy Correct/Assigned	Correct/Attempted	Mean Percent Attempted	Test Scores
Baseline$_1$	34.3	57.0	59.2	62
Reinforcement$_1$	71.2	80.0	89.0	74
Baseline$_2$	56.4	79.0	69.0	67
Reinforcement$_2$	65.5	79.7	82.1	78

It can be seen that all measures were low in baseline$_1$ and then increased in reinforcement$_1$. Then, with the exception of accuracy based on items attempted, all measures decreased in baseline$_2$. Accuracy based on items attempted showed virtually no change. When the contingencies were reinstated in reinforcement$_2$, accuracy based on items attempted maintained its high level while the other measures

returned to their reinforcement$_1$ levels. Again, as in Experiment I, there was a 20 percent reduction in the number of items attempted during baseline$_2$ compared to reinforcement$_1$. However, as was also shown in Experiment I, those items that were attempted in baseline$_2$ were done far more accurately (a gain of 22 percent) than those attempted in baseline$_1$. In addition, as in Experiment I, 10 percent more items were attempted in baseline$_2$ than in baseline$_1$.

Discussion

That each student performed well on the items attempted during the phases of baseline$_2$ is an important finding. Because the number of items attempted during the reversals remained high relative to the reinforcement$_1$ periods, the accuracy level based on items attempted is a valid indicator of performance.

What becomes obvious is that had the accuracy level based on items attempted not been reported, the data would not appear significantly different from those in other studies in which reversal designs had also been used. However, with the reporting of items attempted data, it is apparent that some behavior developed during reinforcement$_1$ was maintained in the absence of external contingencies during baseline$_2$. There are several possible explanations for this phenomenon.

One is that the feedback of number correct, received by the students during all phases, became sufficient to maintain behavior in baseline$_2$, after the feedback had been paired with external rewards in reinforcement$_1$. This would suggest that perhaps some form of intrinsic motivation had developed and the student was then reinforced for simply doing items correctly.

A second possible explanation is that the students learned to become highly selective. Thus, during baseline$_2$, they mainly attempted those items they were sure they could perform correctly. Because none of the material was new to Vic, this hypothesis is less likely to be true in his case. However, in Experiment II, where all of the material was novel, this explanation is plausible.

A third possible explanation that relates exclusively to Experiment I, is that the main difference between Vic's performance during reinforcement$_1$ and baseline$_2$ was one of rate. Although performance on each item was not timed, it is conceivable that Vic worked more slowly during baseline$_2$ than he had during reinforcement$_1$. This would imply that during the development of academic behavior, correct performance can be maintained in the absence of external contingencies more easily than can speed.

In conclusion, both measures of accuracy can be useful in analyzing data, but they must be carefully distinguished. Hopefully, future research will shed more light on the question of why accuracy based on items attempted is maintained in the absence of external rewards.

References

Bushell, D., Wrobel, P., & Michaelis, M. Applying "group" contingencies to the classroom study behavior of preschool children. Journal of Applied Behavior Analysis, 1968, 1, 55-62.

Conlon, M., Hall, C., & Hanley, E. The effects of a peer correction procedure on the arithmetic accuracy for two elementary school children. In G. Semb (Ed.), Behavior analysis and education. Lawrence, Kansas: University of Kansas, 1972.

Ferritor, D., Buckholdt, D., Hamblin, R., & Smith, L. The noneffects of contingent reinforcement for attending behavior on work accomplished. Journal of Applied Behavior Analysis, 1972, 5, 7-18.

Hall, R. V. Managing behavior. Lawrence, Kansas: H & H Enterprises, 1971.

Hall, R. V., Lund, D., & Jackson, D. Effects of teacher attention on study behavior. Journal of Applied Behavior Analysis, 1968, 1, 1-12.

Kirby, F., & Shields, F. Modification of arithmetic response rate and attending behavior in a seventh-grade student. Journal of Applied Behavior Analysis, 1972, 5, 79-84.

Klein, R. D., & Mechelli, T. The modification of on-task and task completion behaviors in an elementary-school child. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1973. Publication 1973/8.

Lovitt, T., Guppy, T., & Blattner, J. The use of a free-time contingency to increase spelling accuracy. Behavior Research and Therapy, 1969, 7, 151-156.

McLaughlin, T., & Malaby, J. Intrinsic reinforcers in a classroom token economy. Journal of Applied Behavior Analysis, 1972, 5, 263-270.

Sulzer, B., Hunt, S., Ashby, E., Komarski, C., & Krams, M. Increasing rate and percentage correct in reading and spelling in a fifth grade public school class of slow readers by means of token system. In E. Ramp & B. L. Hopkins (Eds.), A new direction for education: Behavior analysis. Vol. I. Lawrence, Kansas: University of Kansas, 1971.